Sankhyā : The Indian Journal of Statistics
2007, Volume 69, Part 4, pp. 780-807

On Joint Completeness: Sampling and Bayesian Versions, and Their Connections

Ernesto San Martín
Pontificia Universidad Católica de Chile, Santiago, Chile
Michel Mouchart
Université catholique de Louvain, Louvain-la-Neuve, Belgium

Abstract

Cramer, Kamps and Schenk (*Statist. Decisions*, 2002) established conditions under which a family of joint distributions of two independent statistics is complete, and related their result with a previous one of Landers and Rogge (*Scand. J. Statist.*, 1976). We first propose, within a sampling theory framework, a modification of Cramer, Kamps and Schenk's (2002) generalization, paying a particular attention to the concept of completeness with respect to a *function* of a parameter. Next, after reviewing Bayesian completeness on the sample space, it is shown that Landers and Rogge's (1976) theorem can be extended to a Bayesian framework. A Bayesian version of Cramer, Kamps and Schenk's (2002) theorem is also provided. These results are illustrated with examples in both a normal and a discrete experiment. Finally, taking advantage of the formal symmetry between parameters and observations in a Bayesian experiment, we show that Landers and Rogge type theorems are useful when analysing Bayesian identifiability of structural models often used for modelling individual data.

AMS (2000) *subject classification.* Primary 62H30, 62-07.
Keywords and phrases. Completeness, variation-free parametrization, Rasch Poisson counts model, identification.

1 Introduction

The classical papers of Basu (1955, 1959, 1964) and of Lehmann and Scheffé (1950, 1955) have shown the importance of the completeness of sufficient statistics in the theory of best unbiased estimation and test procedures; see also Lehmann and Casella (1998). Recently, Cramer et al. (2002)

established conditions under which a family of joint distributions of two independent statistics is complete. These authors not only show the practical relevance of their results through a set of interesting examples, but also relate their result with a previous one of Landers and Rogge (1976).

Broadly speaking, Landers and Rogge (1976) state that two independent complete statistics are also complete in the product measure obtained by considering a variation-free parametrization of the product family (or, equivalently, a Cartesian product of the corresponding parameter spaces). Cramer et al.'s (2002) generalization ensures a similar result without requiring the variation-free parametrization between the corresponding families used to define the product family. This generalization relies, however, on a concept not well defined in a pure sampling theory approach, namely the completeness with respect to a *function* of a parameter.

Taking into account that the completeness with respect to a function of a parameter can be well defined in a Bayesian framework, the objective of this paper is to analyse differences and connections between sampling and Bayesian completeness in the context of Landers and Rogge (1976) type theorems. More specifically, within a sampling theory framework, we first propose a modification of Cramer et al.'s (2002) generalization of Landers and Rogge's (1976) theorem; this is developed in Section 2. Next, after reviewing the concept of completeness in a Bayesian framework, we discuss, in Section 3, its robustness with respect to the prior specification and its relationship with sampling completeness. It is then shown, in Section 4, that Landers and Rogge's (1976) theorem can be extended, and in a sense generalized, to a Bayesian framework. A Bayesian version of Cramer et al.'s (2002) theorem is provided in Section 5. In each one of Sections 4 and 5, the results are illustrated with example in both a normal and a discrete Bayesian experiment.

After comparing the Bayesian and the classical versions of the obtained results, Section 7 addresses why should a Bayesian consider completeness. Taking advantage of the formal symmetry between parameters and observations in a Bayesian experiment, not only *parameter completeness* can be defined in a way similar to that of a Bayesian complete statistic, but also an additional Bayesian version of a Landers and Rogge type theorem can be obtained on the parameter space. Such a theorem is useful for analysing the identifiability of certain hierarchical structural models typically used in the context of individual data. This last aspect is illustrated in the context of the Rasch Poisson counts model (see Jansen and van Dujin, 1992; and Jansen, 1994). Let us finally mention that, in a Bayesian framework, identifiability is relevant since *an identified parameter fully concentrates the underlying updating process*; see Florens and Rolin (1984), and Florens, Mouchart and Rolin (1990).

2 Completeness w.r.t a Parameter, Not w.r.t. a Function of a Parameter

Let $\mathcal{E}_s = \{(S, \mathcal{S}), P^\theta : \theta \in \Theta\}$ be a *(parameterized) statistical experiment*, where $(S, \mathcal{S})$ is a measurable space, the *sample space*, and $\{P^\theta : \theta \in \Theta\}$ is a family of *sampling probabilities* defined on the sample space indexed by a *parameter* θ that belongs to a *parameter space* Θ; see, *e.g.*, Barra (1981) or McCullagh (2002). The parameter space Θ might be a Euclidean as well as a functional space, as is the case in non-parametric models, or a product of both as in semi-parametric models. In the context of $\mathcal{E}_s$, both complete statistics and a complete family of probability distributions are defined as follows (see, *e.g.*, Barndorff-Nielsen, 1978; or Barra, 1981).

Definition 2.1. *A statistic $\mathcal{T} \subset \mathcal{S}$ is p-complete $(1 \leq p \leq \infty)$ if for all $t \in \bigcap_{\theta\in\Theta} L^p(S, \mathcal{T}, P^\theta)$,*

$$E^\theta(t) = 0 \ \ \forall\,\theta \in \Theta \Longrightarrow t = 0 \ \ P^\theta\text{-}a.s. \quad \forall\,\theta \in \Theta,$$

where $L^p(S, \mathcal{T}, P^\theta)$ is the linear space of $\mathcal{T}$-measurable functions that are p-integrable w.r.t. P^θ. The family $\{P^\theta : \theta \in \Theta\}$ is said to be p-complete if the statistic $\mathcal{S}$ is p-complete. When $p = \infty$, an ∞-complete statistic is also called a boundedly complete statistic.

Remark 2.1. In this paper, we rely on the usual convention of identifying a statistic $T : (S, \mathcal{S}) \rightarrow (U, \mathcal{U})$ and its generated σ-field $\mathcal{T} = T^{-1}(\mathcal{U}) \doteq \sigma(T) \subset \mathcal{S}$; see, *e.g.*, Barra (1981), Basu and Pereira (1983), Florens et al. (1990) or San Martín, Mouchart and Rolin (2005).

2.1. Joint completeness under non variation-free parametrization. The set-up considered by Cramer et al. (2002) is the following. Let T_1 and T_2 be independent real-valued statistics, and let the induced families of distributions be given by

C0. $$\{P_{T_1}^{\theta_1,\theta_2}\}_{(\theta_1,\theta_2)\in\Theta_1\times\Theta_2}, \quad \text{and} \quad \{P_{T_2}^{\theta_2}\}_{\theta_2\in\Theta_2},$$

i.e., the distribution of T_1 may depend on both parameters, whereas the distribution of T_2 depends on the parameter θ_2 only. Theorem 2 of Cramer et al. (2002) establishes that the family of joint distributions $\{P_{T_1,T_2}^{\theta_1,\theta_2}\}_{(\theta_1,\theta_2)\in\Theta_1\times\Theta_2}$ is complete for $(\theta_1,\theta_2)\in\Theta_1\times\Theta_2$ under the following conditions.

C1. T_1 is complete for θ_1.

C2. T_2 is complete for θ_2.

C3. For all $\theta_1 \in \Theta_1$, $P_{T_1}^{\theta_1,\theta_2{}'}$ and $P_{T_1}^{\theta_1,\theta_2}$ have the same null sets for all θ_2, $\theta_2{}' \in \Theta_2$.

In their introduction, Cramer et al. (2002) recall the definition of a complete statistic, identical to Definition 2.1 above for the case $p = 1$, but fail to define the concept of completeness relative to a *function* of the parameters, such as $f(\theta_1,\theta_2) = \theta_1$, although the use of such a concept is made in condition C1. To the best of these authors' knowledge, such a concept has not been introduced in the statistical literature following a sampling theory approach. We accordingly examine the role of condition C1 in Cramer et al. (2002) result. Let us consider the 1-completeness of T_1 relative to its family of probability distributions indexed by $\theta = (\theta_1,\theta_2)\in\Theta_1\times\Theta_2$, namely that for all $t_1 \in \bigcap_{\theta\in\Theta_1\times\Theta_2} L^1(\mathbb{R},\mathcal{B}_{\mathbb{R}},P_{T_1}^{\theta_1,\theta_2})$, the following implication holds.

$$\begin{aligned} E_{T_1}^{\theta_1,\theta_2}(t_1) &= 0 \quad \forall\,(\theta_1,\theta_2)\in\Theta_1\times\Theta_2 \\ \Longrightarrow \quad t_1 &= 0 \quad P_{T_1}^{\theta_1,\theta_2}\text{-a.s.} \quad \forall\,(\theta_1,\theta_2)\in\Theta_1\times\Theta_2, \end{aligned} \tag{2.1}$$

where $E_{T_1}^{\theta_1,\theta_2}[\cdot]$ denotes the expectation w.r.t. $P_{T_1}^{\theta_1,\theta_2}$. A review of Cramer et al.'s (2002) proof of Theorem 2 leads to the conclusion that condition (2.1) is actually used rather than the undefined condition C1. As a matter of fact, let $g \in L^1(\mathbb{R}^2,\mathcal{B}_{\mathbb{R}^2},P_{T_1,T_2}^{\theta_1,\theta_2})$ be such that $E^{\theta_1,\theta_2}[g(T_1,T_2)] = 0$ for all $(\theta_1,\theta_2)\in\Theta_1\times\Theta_2$, where $E^{\theta_1,\theta_2}[\cdot]$ denotes the expectation taken w.r.t. $P_{T_1,T_2}^{\theta_1,\theta_2}$. Using the same arguments as in Cramer et al. (2002), the independence of T_1 and T_2 implies that $E^{\theta_1,\theta_2}[g(T_1,T_2)] = 0$ is equivalent to

$$\int_{\mathbb{R}} H_{\theta_2}(t_1)dP_{T_1}^{\theta_1,\theta_2}(t_1) = 0 \qquad \forall\,(\theta_1,\theta_2)\in\Theta_1\times\Theta_2, \tag{2.2}$$

where $H_{\theta_2}(t_1) = \int_{\mathbb{R}} g(t_1,t_2)\,dP_{T_2}^{\theta_2}(t_2)$ for all $\theta_2\in\Theta_2$. But the fact that $g \in L^1(\mathbb{R}^2,\mathcal{B}_{\mathbb{R}^2},P_{T_1,T_2}^{\theta_1,\theta_2})$ implies that $H_{\theta_2}(\cdot) \in L^1(\mathbb{R},\mathcal{B},P_{T_1}^{\theta_1,\theta_2})$ $\forall\,(\theta_1,\theta_2)\in\Theta_1\times\Theta_2$. Therefore, conditions (2.1) and (2.2) imply that

$$H_{\theta_2} = 0 \qquad P_{T_1}^{\theta_1,\theta_2}\text{-a.s} \quad \forall\,(\theta_1,\theta_2)\in\Theta_1\times\Theta_2. \tag{2.3}$$

Under C3, condition (2.3) implies that $\int_{\mathbb{R}} g(t_1,t_2)\,dP_{T_2}^{\theta_2}(t_2) = 0$ $P_{T_1}^{\theta_1,\theta_2'}$-a.s for all $\theta_1 \in \Theta_1$ and for all $\theta_2, \theta_2' \in \Theta_2$. The variation-free property between θ_1 and θ_2 in the family of distributions $\{P_{T_1}^{\theta_1,\theta_2} : (\theta_1,\theta_2) \in \Theta_1 \times \Theta_2\}$, (i.e., the Cartesian product structure for the parameter space; see Barndorff-Nielsen, 1978) ensures the validity of the preceding implications. The rest of the proof is as published in Cramer et al. (2002). These arguments motivate to restate Theorem 2 in Cramer et al. (2002) as follows.

Theorem 2.1. *Let T_1 and T_2 be independent statistics satisfying conditions C0, C2, C3 and (2.1). Then the family of joint distributions $\{P_{T_1,T_2}^{\theta_1,\theta_2}\}_{(\theta_1,\theta_2)\in\Theta_1\times\Theta_2}$ is complete for $(\theta_1,\theta_2) \in \Theta_1 \times \Theta_2$.*

Example 2.1. Let us illustrate the use of condition (2.1) with the Example 4 of Cramer et al. (2002). Let T_1 be distributed as a mixture of a uniform distribution on $(-\theta_2, 0)$ with a one-parameter translated exponential distribution on (θ_1, ∞), where $\Theta_1 = \Theta_2 = (0,\infty)$. The corresponding density function is given by

$$f_{T_1}^{\theta_1,\theta_2}(t) = \frac{1}{2}\frac{1}{\theta_2}\mathbb{1}_{[-\theta_2,0]}(t) + \frac{1}{2}e^{-(t_1-\theta_1)}\mathbb{1}_{[\theta_1,\infty)}(t), \quad t \in \mathbb{R},\ (\theta_1,\theta_2) \in \Theta_1 \times \Theta_2.$$

Cramer et al.'s (2002) argument actually shows that T_1 is complete for $(\theta_1,\theta_2) \in \Theta_1 \times \Theta_2$. As a matter of fact, let $g \in L^1(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, P_{T_1}^{\theta_1,\theta_2})$ such that $E^{\theta_1,\theta_2}[g] = 0$ for all $(\theta_1,\theta_2) \in \Theta_1 \times \Theta_2$; it follows that

$$\frac{1}{2}\int_{\theta_1}^{\infty} (c(\theta_2) + g(t))\, e^{-(t-\theta_1)}\, dt \;=\; 0 \quad \forall\, \theta_1 \in \Theta_1 \quad \forall\, \theta_2 \in \Theta_2, \tag{2.4}$$

where $c(\theta_2) = \frac{1}{\theta_2}\int_{-\theta_2}^{0} g(t)\,dt$. By taking an arbitrary but fixed $\theta_2 \in \Theta_2$, equality (2.4) is valid for all $\theta_1 \in \Theta_1$. Therefore, by the completeness of $\{\mathrm{Exp}\,(\theta_1, 1) : \theta_1 \in \Theta_1\}$, (2.4) implies that $g(t) = -c(\theta_2)$ for almost all t. But $E^{\theta_1,\theta_2}[g] = 0$ for all $(\theta_1,\theta_2) \in \Theta_1 \times \Theta_2$; so $c(\theta_2) = 0$ for all $\theta_2 \in \Theta_2$, hence $g = 0$ with respect to the Lebesgue measure.

2.2. Joint completeness under variation-free parametrization. Landers and Rogge (1976) state a different result on the completeness of the family of joint distributions as given below.

Theorem 2.2. (Landers and Rogge, 1976) *Let T_1 and T_2 be independent statistics such that the induced families of distributions have the form*

$\{P_{T_i}^{\theta_i}\}_{\theta_i \in \Theta_i}$ $i = 1, 2$, respectively, and satisfy conditions C1 and C2 above. Then the family of joint distributions $\{P_{T_1,T_2}^{\theta_1,\theta_2}\}_{(\theta_1,\theta_2)\in\Theta_1\times\Theta_2}$ is complete for $\theta = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$.

Theorem 2.2 actually ensures that an arbitrary family of independent and complete statistics is also complete in the product measure obtained by considering a variation-free parametrization of the product family, namely $(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$. Theorem 2.1 is different in nature and ensures that two independent and complete statistics are also complete in the product measure *without* requiring a variation-free parametrization of the corresponding families, (i.e., the parameters of the induced family of T_1 *include* the parameters of the induced family of T_2), but under an additional condition of homogeneity of supports (i.e., condition C3). Let us conclude this section by stating the converse of Theorems 2.1 and 2.2, the proofs of which are straightforward.

THEOREM 2.3.

I. (Converse of Landers and Rogge, 1976) *Let T_1 and T_2 be two independent statistics such that the induced families of distributions have the form $\{P_{T_i}^{\theta_i}\}_{\theta_i\in\Theta_i}$, $i = 1, 2$. If the product family $\{P_{T_1,T_2}^{\theta_1,\theta_2}\}_{(\theta_1,\theta_2)\in\Theta_1\times\Theta_2}$ is complete for $\theta = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$, then each of the families $\{P_{T_i}^{\theta_i}\}_{\theta_i\in\Theta_i}$ $i = 1, 2$ satisfies conditions C1 and C2 above.*

II. (Converse of Theorem 2.1) *Let T_1 and T_2 be two independent statistics satisfying condition C0 above. If the product family $\{P_{T_1,T_2}^{\theta_1,\theta_2}\}_{(\theta_1,\theta_2)\in\Theta_1\times\Theta_2}$ is complete for $\theta = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$, then each of the families $\{P_{T_1}^{\theta_1,\theta_2}\}_{(\theta_1,\theta_2)\in\Theta_1\times\Theta_2}$ and $\{P_{T_2}^{\theta_2}\}_{\theta_2\in\Theta_2}$ satisfies conditions C2 and (2.1) above.*

3 Bayesian Completeness on the Sampling Space

In Section 2, we saw that the main result of Cramer et al. (2002) does not require a (new) concept of completeness relative to a function of parameters because its proof only uses the standard concept of completeness. An issue is to understand why such a concept has not been developed in the sampling theory approach in spite of the fact that many (or most) statistical models involve nuisance parameters, making the parameter of interest a non-injective function of the parameter θ. This may be due to the fact that sampling theory does not provide a general procedure for eliminating nuisance parameters (see, *e.g.*, Basu, 1977; and Berti et al., 2000), which is at variance with Bayesian theory, where nuisance parameters are integrated out by means of the (conditional) prior distribution. In order to make this paper reasonably self-contained, let us review the Bayesian concept of completeness.

3.1. A Bayesian experiment. Consider the statistical experiment $\mathcal{E}_s$ as defined in Section 2. A probability measure Q on $\Theta \times S$ is constructed by endowing the parameter space Θ with a probability measure m on $(\Theta, \mathcal{A})$, where the σ-field $\mathcal{A}$ of subsets of Θ makes $P^\theta(X)$ measurable for all $X \in \mathcal{S}$, and by extending the function Q defined on $\mathcal{A} \times \mathcal{S}$ to $\mathcal{A} \otimes \mathcal{S}$ (in a unique way) as follows.

$$Q(E \times X) \;=\; \int_E P^\theta(X)\, dm \qquad E \in \mathcal{A}, \quad X \in \mathcal{S}. \tag{3.1}$$

The measure constructed in (3.1) is denoted as $Q = m \otimes P^{\mathcal{A}}$. Thus, a *Bayesian experiment* is defined as the probability space $\mathcal{E} = (\Theta \times S, \mathcal{A} \vee \mathcal{S}, Q = m \otimes P^{\mathcal{A}})$.

REMARK 3.1. Similar to Remark 2.1, we systematically identify the sub-σ-field $\mathcal{B} \subset \mathcal{A}$ (resp., $\mathcal{T} \subset \mathcal{S}$) with the sub-$\sigma$-field of the corresponding cylinders $\mathcal{B} \times S$ (resp. $\Theta \times \mathcal{T}$). Thus, in the Bayesian experiment $\mathcal{E}$, we identify the product $\mathcal{A} \otimes \mathcal{S}$ with $\mathcal{A} \vee \mathcal{S}$, the σ-field generated by $(\mathcal{A} \times S) \cup (\Theta \times \mathcal{S})$.

By construction, P^θ in (3.1) becomes a transition probability representing a regular version of $P^{\mathcal{A}}$, the restriction to $\mathcal{S}$ of the conditional probability Q given $\mathcal{A}$, and this is so for any probability m on $(\Theta, \mathcal{A})$. Moreover, the so-called *prior probability* m corresponds to the marginal probability of Q on $(\Theta, \mathcal{A})$, namely $m(E) = Q(E \times S)$ for $E \in \mathcal{A}$. Similarly, the marginal probability P on the sample space $(S, \mathcal{S})$ given by $P(X) = Q(\Theta \times X)$ for $X \in \mathcal{S}$ is called the *predictive probability.*

Apart from the decomposition $Q = m \otimes P^{\mathcal{A}}$, the probability Q is decomposed into a marginal probability P on $(S, \mathcal{S})$ and, under the usual hypotheses, a regular conditional probability given $\mathcal{S}$, represented by a transition denoted as $m^{\mathcal{S}}$: this is the so-called *posterior distribution.* When Q is decomposed as $Q = m \otimes P^{\mathcal{A}} = P \otimes m^{\mathcal{S}}$, the Bayesian experiment $\mathcal{E}$ is said to be *regular.* For more details, see, *e.g.*, Martin et al. (1973) and Florens et al. (1990, Chapter 1).

3.2. Bayesian completeness and its relation with sampling completeness. In the context of the Bayesian experiment $\mathcal{E}$, the sub-σ-field $\mathcal{T}$ of $\mathcal{S}$ corresponds to a statistic, whereas the sub-σ-field $\mathcal{B}$ of $\mathcal{A}$ corresponds to a function of parameters; see Remark 2.1. The completeness of a statistic with respect to a parameter is defined, both in the marginal and in the conditional case, as follows.

DEFINITION 3.1. *A statistic $\mathcal{T} \subset \mathcal{S}$ is p-complete $(1 \leq p \leq \infty)$ with respect to a parameter $\mathcal{B} \subset \mathcal{A}$ if the following implication holds:*

$$\forall\, t \in L^p(\Theta \times S, \mathcal{T}, Q_{\mathcal{B}\vee\mathcal{T}}), \qquad E(t \mid \mathcal{B}) = 0 \Longrightarrow t = 0 \qquad m_{\mathcal{B}}\text{-a.s.}, \quad (3.2)$$

where $m_{\mathcal{B}}$ is the trace, on $\mathcal{B}$, of the prior probability m. We denote it as $\mathcal{T} \ll_p \mathcal{B}$. Thus, completeness corresponds to the injectivity of the sampling expectation, as an operator defined on the (integrable) functions on the sample space. Let $\mathcal{M} \subset \mathcal{A} \vee \mathcal{S}$ be a sub-σ-field. Conditionally on $\mathcal{M}$, a statistic $\mathcal{T} \subset \mathcal{S}$ is p-complete $(1 \leq p \leq \infty)$ w.r.t. a parameter $\mathcal{B} \subset \mathcal{A}$ if $\mathcal{T} \vee \mathcal{M}$ is p-complete $(1 \leq p \leq \infty)$ w.r.t. $\mathcal{B} \vee \mathcal{M}$. We denote it as $\mathcal{T} \ll_p \mathcal{B} \mid \mathcal{M}$.

In this definition, the σ-field $\mathcal{M}$ can be a parameter, or a statistic, or a function of both. For properties, see Basu and Pereira (1983), Mouchart and Rolin (1984), and Florens et al. (1990, Chapter 5).

The relationships between Bayesian completeness and sampling completeness essentially depend on the regularity of the prior specification. We say that the prior probability m is *regular for the statistical experiment $\mathcal{E}_s$* if for a bounded $\mathcal{S}$-measurable function s such that $E^\theta(s) = 0$ m-a.s., it follows that $E^\theta(s) = 0$ for all $\theta \in \Theta$. Two relevant cases of regularity are the following: (i) if Θ is countable, the prior probability m is regular if it gives positive mass to each point of Θ; (ii) if Θ is a topological space, and the sampling probabilities are such that P^θ is continuous on Θ, then a prior probability m is regular if it gives positive probability to each nonempty open measurable subset of Θ. The following theorem relates Bayesian and sampling completeness; for a proof, see Florens et al. (1990, Section 5.5.4).

THEOREM 3.1. *Let us consider the statistical experiment $\mathcal{E}_s$ and the Bayesian experiment $\mathcal{E}$. A statistic, which is p-complete with respect to $\mathcal{A}$ in the context of $\mathcal{E}$ characterized by $Q = m \otimes P^{\mathcal{A}}$, will be sampling p-complete, and conversely, once the prior probability m is regular.*

3.3. Robustness with respect to the prior specification. In (3.2), the completeness of a statistic $\mathcal{T}$ with respect to a parameter $\mathcal{B}$ depends on the prior specification in two ways. Let us decompose m on $\mathcal{A}$ with respect to $\mathcal{B} \subset \mathcal{A}$, namely $m = m_{\mathcal{B}} \otimes m^{\mathcal{B}}$, where $m^{\mathcal{B}}$ is a conditional probability of m given $\mathcal{B}$. We assume the existence of a regular version of $m^{\mathcal{B}}$, which enters in the construction of $E(t \mid \mathcal{B})$ because

$$E(t \mid \mathcal{B}) = \int t \, dP^{\mathcal{B}}, \qquad \text{where} \qquad P^{\mathcal{B}}(X) = \int_{\Theta} P^{\mathcal{A}}(X) \, dm^{\mathcal{B}}, \quad X \in \mathcal{S}.$$

Next, $m_{\mathcal{B}}$ determines the null sets describing the almost-sure equality $E(t \mid \mathcal{B}) = 0$. Therefore, this completeness is robust to a modification of the prior distribution leaving $m^{\mathcal{B}}$ unchanged and leaving the collection of null sets of $m_{\mathcal{B}}$ unaffected. Thus, when $\mathcal{B} = \mathcal{A}$, the validity of $E(t \mid \mathcal{A}) = 0$ depends only on the null sets of m, and accordingly condition (3.2) is robust to any equivalent modification of the prior specification. So, we have given a simple proof of the following theorem.

THEOREM 3.2. *If $\mathcal{T} \ll_p \mathcal{A}$ $(1 \leq p \leq \infty)$ in the Bayesian experiment characterized by $Q = m \otimes P^{\mathcal{A}}$, then $\mathcal{T} \ll_p \mathcal{A}$ for all $Q' = m' \otimes P^{\mathcal{A}}$ once m and m' have the same null sets.*

4 A Bayesian Version of Landers and Rogge's Theorem

The objective of this section is to extend Landers and Rogge (1976) theorem to a Bayesian framework. The tool of conditional independence is needed. Although well known, let us briefly recall the definition. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{F}_i$, with $i = 1, 2, 3$, be sub-σ-fields of $\mathcal{F}$. Then $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2 \mid \mathcal{F}_3$ if and only if $E[f \mid \mathcal{F}_2 \vee \mathcal{F}_3] = E[f \mid \mathcal{F}_3]$ for all $\mathcal{F}_1$-measurable and bounded function f or, equivalently, $E[f_1 f_2 \mid \mathcal{F}_3] = E[f_1 \mid \mathcal{F}_3] \cdot E[f_2 \mid \mathcal{F}_3]$ for all $\mathcal{F}_i$-measurable and bounded function f_i with $i = 1, 2$. For details, proofs and properties, see, *e.g.*, Martin et al. (1973), Dawid (1980), Döhler (1980), Mouchart and Rolin (1984), or Florens et al. (1990, Chapter 2).

4.1. Main result. In what follows, we use the following definition.

DEFINITION 4.1. *Let $\mathcal{E} = (\Theta \times S, \mathcal{A} \vee \mathcal{S}, Q)$ be a Bayesian experiment as introduced in Section 3.1. Let $\mathcal{B} \subset \mathcal{A}$ be a parameter and $\mathcal{T} \subset \mathcal{S}$ be a statistic. If $\mathcal{T} \perp\!\!\!\perp \mathcal{A} \mid \mathcal{B}$, it is said that $\mathcal{B}$ is a sufficient parameter for $\mathcal{T}$. If $\mathcal{T} \perp\!\!\!\perp \mathcal{B}$, it is said that $\mathcal{B}$ and $\mathcal{T}$ are mutually ancillary.*

REMARK 4.1. The condition $\mathcal{T} \perp\!\!\!\perp \mathcal{A} \mid \mathcal{B}$ means that the process generating $\mathcal{T}$ given $\mathcal{A}$ depends on $\mathcal{B}$ only; therefore, the parameter $\mathcal{B}$ is "sufficient" to describe the sampling process generating $\mathcal{T}$, $\mathcal{A}$ being redundant. In particular, $\mathcal{A} \perp\!\!\!\perp \mathcal{S} \mid \mathcal{B}$ means that $\mathcal{B}$ is a sufficient parameter for the complete data $\mathcal{S}$: this is the same concept as $\mathcal{A} \perp\!\!\!\perp \mathcal{S} \mid \mathcal{T}$ for $\mathcal{T}$ being a sufficient statistic for the complete parameter $\mathcal{A}$. The condition $\mathcal{T} \perp\!\!\!\perp \mathcal{B}$ means that the parameter $\mathcal{B}$ (resp. the statistic $\mathcal{T}$) is of no actual relevance for describing the process generating $\mathcal{T}$ (resp. the process generating $\mathcal{B}$). For more details, see Florens et al. (1990, Chapters 2 and 3).

The following theorem corresponds to a Bayesian version of Landers and Rogge's (1976) theorem. The proof is given in the Appendix.

THEOREM 4.1. *Let* $(\mathcal{T}_i, \mathcal{B}_i)$ *with* $i = 1, 2$ *be two pairs of statistics and parameters such that*

$$\textit{(i)} \quad \mathcal{T}_1 \perp\!\!\!\perp \mathcal{B}_2 \mid \mathcal{B}_1 \qquad \textit{and} \quad \textit{(ii)} \quad \mathcal{T}_2 \perp\!\!\!\perp \mathcal{B}_1 \mid \mathcal{B}_2. \tag{4.1}$$

For all $1 \leq p \leq \infty$,

I. (Bayesian version of Landers and Rogge, 1976) *If* $\mathcal{T}_i \ll_p \mathcal{B}_i$ *with* $i = 1, 2$*, and if*

$$\textit{(i)}\ \mathcal{T}_1 \perp\!\!\!\perp \mathcal{T}_2 \mid \mathcal{B}_1 \vee \mathcal{B}_2, \qquad \textit{and} \quad \textit{(ii)}\ \mathcal{B}_1 \perp\!\!\!\perp \mathcal{B}_2, \tag{4.2}$$

then $(\mathcal{T}_1 \vee \mathcal{T}_2) \ll_p (\mathcal{B}_1 \vee \mathcal{B}_2)$.

II. (Converse version) *If* $(\mathcal{T}_1 \vee \mathcal{T}_2) \ll_p (\mathcal{B}_1 \vee \mathcal{B}_2)$*, then* $\mathcal{T}_i \ll_p \mathcal{B}_i$ *with* $i = 1, 2$.

The following are some remarks on the hypotheses and the conclusions of this theorem.

REMARK 4.2. Condition (4.1) defines $\mathcal{B}_i$ as a sufficient parameter for $\mathcal{T}_i$, for $i = 1, 2$; see Definition 4.1. In a pure sampling theory approach, this condition corresponds to the property that the distribution of the statistic $\mathcal{T}_i$ depends on the parameter $\mathcal{B}_i$ only, with $i = 1, 2$.

REMARK 4.3. The condition (4.2)(i) is the Bayesian counterpart of the sampling independence between $\mathcal{T}_1$ and $\mathcal{T}_2$.

Remark 4.4. The condition (4.2)(ii) of prior independence between $\mathcal{B}_1$ and $\mathcal{B}_2$ is the Bayesian counterpart of the variation-free property between the corresponding parameter spaces in a pure sampling approach. This condition is needed to establish the implication; if the condition does not hold, in particular if $\mathcal{B}_1 = \mathcal{B}_2$, the theorem is not valid any more because $\mathcal{B}$ cannot be independent of itself, except in the trivial case of a known parameter.

Example 4.1. As a simple example of the non-validity of part I of Theorem 4.1, when $\mathcal{B}_1 = \mathcal{B}_2$, consider two independent samples from a $\mathcal{N}(\theta, 1)$ and a regular prior distribution m giving positive probability to each nonempty open measurable subset of $\mathbb{R}$. It follows that both $\bar{X}_1$ and $\bar{X}_2$ are complete w.r.t. θ. Nevertheless, $\sigma(\bar{X}_1 - \bar{X}_2) \subset \sigma(\bar{X}_1, \bar{X}_2)$ is not complete w.r.t. θ since $E[\bar{X}_1 - \bar{X}_2 \mid \theta] = 0$.

Remark 4.5. It should be noticed that the two conditions in (4.1), along with the two conditions (4.2), jointly imply the following conditions:

$$\begin{array}{llllll} \text{(i)} & \mathcal{T}_1 \perp\!\!\!\perp \mathcal{B}_2 \mid \mathcal{T}_2, & \text{(ii)} & \mathcal{T}_2 \perp\!\!\!\perp \mathcal{B}_1 \mid \mathcal{T}_1, & \text{(iii)} & \mathcal{T}_1 \perp\!\!\!\perp \mathcal{T}_2, \\ \text{(iv)} & \mathcal{B}_1 \perp\!\!\!\perp \mathcal{B}_2 \mid \mathcal{T}_1 \vee \mathcal{T}_2, & \text{(v)} & \mathcal{T}_1 \perp\!\!\!\perp \mathcal{B}_2, \text{ and} & \text{(vi)} & \mathcal{T}_2 \perp\!\!\!\perp \mathcal{B}_1. \end{array} \tag{4.3}$$

Condition (4.3)(i) means that $\mathcal{T}_2$ is a sufficient statistic for $\mathcal{B}_2$ (after integrating out $\mathcal{B}_1$), whereas Condition (4.3)(ii) means that $\mathcal{T}_1$ is a sufficient statistic for $\mathcal{B}_1$ (after integrating out $\mathcal{B}_2$). Condition (4.3)(iii) means that $\mathcal{T}_1$ and $\mathcal{T}_2$ are predictively independent, whereas condition (4.3)(iv) means that $\mathcal{B}_1$ and $\mathcal{B}_2$ are *a posteriori* mutually independent. Finally, condition (4.3)(v) (resp. condition (4.3)(vi)) means that $\mathcal{T}_1$ and $\mathcal{B}_2$ (resp. $\mathcal{T}_2$ and $\mathcal{B}_1$) are mutually ancillary; see Definition 4.1.

Remark 4.6. The first part of Theorem 4.1 is a Bayesian analogue of Landers and Rogge (1976), where the variation-free condition becomes a condition of prior independence. For the converse part, the Bayesian version requires neither the sampling independence nor the prior independence, which is at variance with the sampling version in part I of Theorem 2.3.

4.2. Application to a normal conjugate Bayesian experiment. According to Theorem 3.1, if regular prior distributions are considered on the parameters of the distributions used to illustrate Landers and Rogge's (1976) theorem, the same examples are still valid as illustrations of part I of Theorem 4.1. In this section, we illustrate Theorem 4.1 when the prior probability distribution is *not* regular, which leads to consideration of non-trivial prior null sets; for details about null sets, see San Martín et al. (2005).

Let $X = ({X_1}', {X_2}', {X_3}')' \in \mathbb{R}^{p_1+p_2+p_3}$ be a random vector. Let $V(\cdot \mid \cdot)$ and $C(\cdot, \cdot \mid \cdot)$ denote the conditional variance and the conditional covariance operators, respectively, and define

$$\begin{aligned} \text{Ker}\,[C(X_2, X_1 \mid X_3)] &= \{a \in \mathbb{R}^{p_1} : C(X_2, a'X_1 \mid X_3) = 0 \quad \text{a.s.}\} \text{ and} \\ \text{Ker}\,[V(X_1 \mid X_3)] &= \{a \in \mathbb{R}^{p_1} : V(a'X_1 \mid X_3) = 0 \quad \text{a.s.}\}. \end{aligned}$$

The following proposition characterizes the p-completeness of X_1 w.r.t. X_2 conditionally on X_3 whether the covariance matrix is singular or not; for a proof, see Appendix.

PROPOSITION 4.1. *Let* $({X_1}', {X_2}' \mid {X_3}')' \sim \mathcal{N}_{p_1+p_2}(\mu(X_3), \Sigma(X_3))$. *The following statements are* X_3*-a.s. equivalent.*

(i) $\forall\, p \in [1, \infty]$, $X_1 \ll_p X_2 \mid X_3$.

(ii) $r[C(X_2, X_1 \mid X_3)] = r[V(X_1 \mid X_3)]$.

(iii) $Ker[C(X_2, X_1 \mid X_3)] = Ker[V(X_1 \mid X_3)]$.

(iv) $Ker[C(X_2, X_1 \mid X_3)] \subset Ker[V(X_1 \mid X_2, X_3)]$.

If X_1 is the observed information T, X_2 is the parameter θ, and X_3 is a.s. a constant, then the following can be concluded from Proposition 4.1.

1. When the covariance matrix of $(T', \theta')'$ is positive definite, $T \ll_p \theta$ *if and only if* the number of parameters is at most equal to the number of observations. In this case, the prior distribution on θ is regular since $r[V(\theta)] = p_2$, and accordingly this p-complete relationship is also valid in a pure sampling framework.

2. When the covariance matrix of $(T', \theta')'$ is singular, *a necessary, but not a sufficient condition* to ensure the p-completeness of T w.r.t. θ, is that the number of parameters be at most equal to $r[V(T)]$.

EXAMPLE 4.2. Let $T = ({T_1}', {T_2}')' \in \mathbb{R}^{p_1+p_2}$ be a manifest variable analysed under a random effect model parameterized by $\theta = ({\theta_1}', {\theta_2}')' \in \mathbb{R}^{q_1+q_2}$. We only consider, without making this explicit, joint distributions of (T, θ) conditional on their expectation, assumed equal to 0, and on their covariance matrix, namely $(T', \theta')' \sim \mathcal{N}_{p+q}(0, \Sigma)$, with $p = p_1 + p_2$ and $q = q_1 + q_2$, where Σ might be singular.

Example 4.2 related to Theorem 4.1, part I. Assume that $T_i \ll_p \theta_i$, with $i = 1, 2$. Using Proposition 4.1, these conditions are equivalent to

$$\text{(i) } r[V(T_1)] = r[C(\theta_1, T_1)], \quad \text{and} \quad \text{(ii) } r[V(T_2)] = r[C(\theta_2, T_2)]. \tag{4.4}$$

Equation (4.3)(iii) means that $V(T) = \text{diag}[V(T_1), V(T_2)]$, namely a block-diagonal matrix, with the matrices $V(T_1)$ and $V(T_2)$ as the corresponding blocks. Similarly, from (4.3)(v-vi), it follows that

$$C[\theta, T] = \text{diag}\,[C(\theta_1, T_1), C(\theta_2, T_2)].$$

Taking into account this block-diagonal structure, condition (4.4) implies in a straightforward way that $r[V(T)] = r[C(\theta, T)]$, which is equivalent to $T \ll_p \theta$. Let us note that the singularity of $V(\theta)$ would mean a linear relation between some elements of the random vector θ but does not play any role in the conclusion, which is at variance from a result similar to the sampling one. Finally, according to Theorem 3.2, the p-completeness of T w.r.t. θ is still valid if the prior distributions on the θ_i's are replaced by equivalent prior ones.

Example 4.2 related to Theorem 4.1, part II. Assume that the pairs (T_i, θ_i), with $i = 1, 2$, satisfy condition (4.1) above; (4.1)(i) is equivalent to $C(T_1, \theta_2 \mid \theta_1) = 0$. Moreover, from the normality distribution, it follows that $E(\theta_2 \mid \theta_1) = C(\theta_1, \theta_2)\,[V(\theta_1)]^+\,\theta_1$ and $E(T_1 \mid \theta_1) = C(T_1, \theta_1)\,[V(\theta_1)]^+\,\theta_1$, where A^+ denotes the Moore-Penrose inverse of A (see Marsaglia, 1964). Using the iterative decomposition of a covariance, along with the fact that $A^+AA^+ = A^+$, it follows that

$$C(\theta, T) \;=\; \begin{pmatrix} I_{q_1} & R_{12} \\ Q_{21} & I_{q_2} \end{pmatrix} \begin{pmatrix} C(\theta_1, T_1) & 0 \\ 0 & C(\theta_2, T_2) \end{pmatrix},$$

where $Q_{21} = C(\theta_2, \theta_1)[V(\theta_1]^+$ and $R_{12} = C(\theta_1, \theta_2)[V(\theta_2)]^+$. Thus, if $T \ll_p \theta$, then by Proposition 4.1,

$$\text{Im}[V(T)] = \text{Im}\begin{bmatrix} C(T_1, \theta_1) \\ 0 \end{bmatrix} \oplus \text{Im}\begin{bmatrix} 0 \\ C(T_2, \theta_2) \end{bmatrix},$$

where $\text{Im}(A)$ denotes the range space generated by the columns of matrix A. It follows that $r[V(T_i)] = r[C(\theta_i, T_i)]$ with $i = 1, 2$, that is, $T_i \ll_p \theta_i$ with

$i = 1, 2$. As mentioned in part II of Theorem 4.1, the conclusion depends neither on the sampling independence between T_1 and T_2, nor on the prior independence between θ_1 and θ_2.

4.3. Application to a discrete Bayesian experiment. Let us characterize Bayesian completeness in the discrete case. Let $(M, \mathcal{M}, P)$ be a probability space, N_r for $r = 1, 2, 3$ be finite sets, and $X_r : M \longrightarrow N_r$ with $r = 1, 2, 3$ be random variables. We define $K = \{k \in N_3 : P[X_3 = k] > 0\}$ and, for each $k \in K$,

$$\begin{aligned} N_1^{(k)} &= \{i \in N_1 : P[X_1 = i \mid X_3 = k] > 0\}, \\ N_2^{(k)} &= \{j \in N_2 : P[X_2 = j \mid X_3 = k] > 0\}; \end{aligned}$$

and the $|N_1^{(k)}| \times |N_2^{(k)}|$ matrix $\mathbf{P}^{(k)}$ with the elements

$$p_{ij|k} \doteq (\mathbf{P}^{(k)})_{ij} = P[X_1 = i,\, X_2 = j \mid X_3 = k] \quad \text{for } (i,j) \in N_1^{(k)} \times N_2^{(k)}.$$

The following proposition characterizes Bayesian completeness in the discrete case; for a proof, see Appendix.

PROPOSITION 4.2. *For $p \geq 1$, $X_1 \ll_p X_2 \mid X_3$ if and only if* $(\forall\, k \in K)$ $\mathbf{P}^{(k)\prime}$ *is an injective linear transformation, i.e.,* $r(\mathbf{P}^{(k)}) = |N_1^{(k)}|$.

REMARK 4.7. If $X_1 \ll_p X_2 \mid X_3$, then a dimension restriction between X_1 and X_2 follows, namely that for each $k \in K$, $r(\mathbf{P}^{(k)}) = |N_1^{(k)}| \leq |N_2^{(k)}|$.

REMARK 4.8. $\mathbf{P}^{(k)}$ is a bijective linear transformation, i.e., $|N_1^{(k)}| = |N_2^{(k)}|$) for each $k \in K$ if and only if $X_1 \ll_p X_2 \mid X_3$ and $X_2 \ll_p X_1 \mid X_3$.

EXAMPLE 4.3. Let $(T_1, T_2, \theta_1, \theta_2) \in \{0, 1\}^4$. Without restrictions, this Bayesian experiment has $2^4 - 1 = 15$ parameters. Let $W = [\omega_{ijkl}]$ be the 4×4 matrix of joint probabilities, where $\omega_{ijkl} \doteq P[T_1 = i, T_2 = j, \theta_1 = k, \theta_2 = l]$.

EXAMPLE 4.3 RELATED TO THEOREM 4.1, PART I. Under conditions (4.1) and (4.2), the joint probability distribution is decomposed as follows.

$$\begin{aligned} \omega_{ijkl} &= P[T_1 = i \mid \theta_1 = k]\, P[\theta_1 = k]\; P[T_2 = j \mid \theta_2 = l]\, P[\theta_2 = l] \\ &\doteq p_{i|k}\, m_k \;\cdot\; q_{j|l}\, n_l. \end{aligned}$$

Therefore, the Bayesian experiment is characterized by 6 parameters only: $p_{0|0}, p_{0|1}, q_{0|0}, q_{0|1}, m_0, n_0$. By Proposition 4.2, the p-completeness of T_1 w.r.t θ_1 requires to analyse the rank of the 2×2 matrix with entries of the form $r_{ik} \doteq P[T_1 = i, \theta_1 = k] = p_{i|k} m_k$, namely

$$\begin{bmatrix} r_{00} & r_{01} \\ r_{10} & r_{11} \end{bmatrix} = \begin{bmatrix} p_{0|0} & p_{0|1} \\ p_{1|0} & p_{1|1} \end{bmatrix} \begin{bmatrix} m_0 & 0 \\ 0 & m_1 \end{bmatrix}. \tag{4.5}$$

Similarly, the p-completeness of T_2 w.r.t. θ_2 requires to analyse the rank of the 2×2 matrix with entries of the form $s_{ik} \doteq P[T_2 = j, \theta_2 = l] = q_{j|l} n_l$, namely

$$\begin{bmatrix} s_{00} & s_{01} \\ s_{10} & s_{11} \end{bmatrix} = \begin{bmatrix} q_{0|0} & q_{0|1} \\ q_{1|0} & q_{1|1} \end{bmatrix} \begin{bmatrix} n_0 & 0 \\ 0 & n_1 \end{bmatrix}. \tag{4.6}$$

Finally, the p-completeness of (T_1, T_2) w.r.t. (θ_1, θ_2) requires to analyse the rank of the 4×4 matrix with entries of the form ω_{ijkl}. It can be easily verified that

$$W = \begin{bmatrix} \omega_{0000} & \omega_{0001} & \omega_{0010} & \omega_{0011} \\ \omega_{0100} & \omega_{0101} & \omega_{0110} & \omega_{0111} \\ \omega_{1000} & \omega_{1001} & \omega_{1010} & \omega_{1011} \\ \omega_{1100} & \omega_{1101} & \omega_{1110} & \omega_{1111} \end{bmatrix} = \begin{bmatrix} r_{00} & r_{01} \\ r_{10} & r_{11} \end{bmatrix} \otimes \begin{bmatrix} s_{00} & s_{01} \\ s_{10} & s_{11} \end{bmatrix}, \tag{4.7}$$

which we write as $W = R \otimes S$. This equality shows in particular the role of both the prior independence of θ_1 and θ_2, and the sampling independence of T_1 and T_2.

Now, $T_1 \ll_p \theta_1$ if and only if $r(R) = 2$ which, by (4.5), is equivalent to both $m_0 m_1 > 0$ and $P[T_1 = 0 \mid \theta_1 = 0] \neq P[T_1 = 0 \mid \theta_1 = 1]$. Similarly, $T_2 \ll_p \theta_2$ if and only if $r(S) = 2$, which, by (4.6), is equivalent to both $n_0 n_1 > 0$ and $P[T_2 = 0 \mid \theta_2 = 0] \neq P[T_2 = 0 \mid \theta_2 = 1]$. Finally, equality (4.7) shows that $r(R) = 2$ and $r(S) = 2$ jointly imply that $r(W) = 4$, which is equivalent to $(T_1, T_2) \ll_p (\theta_1, \theta_2)$.

Example 4.3 related to Theorem 4.1, part II. Assume that the pair (T_i, θ_i) with $i = 1, 2$ satisfies condition (4.1) above. Let us assume that $(T_1, T_2) \ll_p (\theta_1, \theta_2)$. By Proposition 4.2, this is equivalent to $r(W) = 4$. Taking into account condition (4.1)(ii), $T_2 \ll_p \theta_2$ if the 2×2 matrix with entries $P[T_2 = j, \theta_2 = l]$ is a full rank matrix; this last matrix can equivalently be rewritten as

$$\begin{pmatrix} \omega_{\cdot 0 \cdot 0} & \omega_{\cdot 0 \cdot 1} \\ \omega_{\cdot 1 \cdot 0} & \omega_{\cdot 1 \cdot 1} \end{pmatrix} = \begin{pmatrix} \omega_{0000}+\omega_{1000}+\omega_{0010}+\omega_{1010} & \omega_{0001}+\omega_{0011}+\omega_{1001}+\omega_{1011} \\ \omega_{0100}+\omega_{0110}+\omega_{1100}+\omega_{1110} & \omega_{0101}+\omega_{0111}+\omega_{1101}+\omega_{1111} \end{pmatrix}.$$

Assume that the rank of this matrix is equal to 1. Then there exists a constant $c \neq 0$ such that

$$\begin{aligned}\omega_{0000} + \omega_{1000} + \omega_{0010} + \omega_{1010} &= c\,[\omega_{0001} + \omega_{0011} + \omega_{1001} + \omega_{1011}],\\ \omega_{0100} + \omega_{0110} + \omega_{1100} + \omega_{1110} &= c\,[\omega_{0101} + \omega_{0111} + \omega_{1101} + \omega_{1111}].\end{aligned}$$

These conditions imply that the first and third rows of W are linearly dependent, and that the second and fourth rows of W are also linearly dependent. This contradicts the fact that $r(W) = 4$. Therefore, $T_2 \ll_p \theta_2$. Similarly, it can be concluded that $(T_1, T_2) \ll_p (\theta_1, \theta_2)$ implies that $T_1 \ll_p \theta_1$. As mentioned in part II of Theorem 4.1, the conclusion depends neither on the sampling independence between T_1 and T_2, nor on the prior independence between θ_1 and θ_2.

5 Joint Bayesian Completeness without Prior Independence

The objective of this section is to obtain a Bayesian version of Theorem 2.1 and to provide an illustration of it. Except condition C3, the other hypotheses underlying Theorem 2.1 have an obvious Bayesian counterpart. Note first that a converse Bayesian version of Theorem 2.1 in the same line as part II of Theorem 4.1, is trivially obtained by formally replacing $\mathcal{B}_1$ by $\mathcal{B}_1 \vee \mathcal{B}_2$ in conditions (4.1), in which case condition (4.1)(i) becomes trivial. Thus, $\mathcal{B}_1 \vee \mathcal{B}_2$ is a sufficient parameter for $\mathcal{T}_1$, whereas $\mathcal{B}_2$ is a sufficient parameter for $\mathcal{T}_2$; see Definition 4.1.

THEOREM 5.1. (Bayesian Converse version of Theorem 2.1) *Let $p \in [1, \infty]$ and let $(\mathcal{T}_i, \mathcal{B}_i)$ with $i = 1, 2$ be two pairs of statistics and parameters such that $\mathcal{B}_1 \vee \mathcal{B}_2$ is sufficient for $\mathcal{T}_1$, and $\mathcal{B}_2$ is sufficient for $\mathcal{T}_2$. If $(\mathcal{T}_1 \vee \mathcal{T}_2) \ll_p (\mathcal{B}_1 \vee \mathcal{B}_2)$, then $\mathcal{T}_1 \ll_p (\mathcal{B}_1 \vee \mathcal{B}_2)$ and $\mathcal{T}_2 \ll_p \mathcal{B}_2$.*

Let us now consider the following question: which conditions should be added to both conditions (4.1)(ii) and (4.2) to ensure that $(\mathcal{T}_1 \vee \mathcal{T}_2) \ll_p (\mathcal{B}_1 \vee \mathcal{B}_2)$ with $p \in [1, \infty]$? Let $t \in L^p(\Theta \times S, \mathcal{T}_1 \vee \mathcal{T}_2, Q_{\mathcal{B}_1 \vee \mathcal{B}_2 \vee \mathcal{T}_1 \vee \mathcal{T}_2})$ such that $E(t \mid \mathcal{B}_1 \vee \mathcal{B}_2) = 0$. Conditions (4.1)(ii) and (4.2)(i) are jointly equivalent to $\mathcal{T}_2 \perp\!\!\!\perp (\mathcal{B}_1 \vee \mathcal{T}_1) \mid \mathcal{B}_2$, which implies that $\mathcal{T}_1 \vee \mathcal{T}_2 \perp\!\!\!\perp \mathcal{B}_1 \mid \mathcal{T}_1 \vee \mathcal{B}_2$. Thus, $E(t \mid \mathcal{B}_1 \vee \mathcal{B}_2 \vee \mathcal{T}_1) = E(t \mid \mathcal{B}_2 \vee \mathcal{T}_1)$; and, consequently, $0 = E(t \mid \mathcal{B}_1 \vee \mathcal{B}_2) = E[E(t \mid \mathcal{B}_2 \vee \mathcal{T}_1) \mid \mathcal{B}_1 \vee \mathcal{B}_2]$. Therefore, if *conditionally on $\mathcal{B}_2$, $\mathcal{T}_1$ is p-complete w.r.t. $\mathcal{B}_1$*, it follows that $E(t \mid \mathcal{B}_2 \vee \mathcal{T}_1) = 0$. This last equality implies that $t = 0$ a.s. under the condition that, *conditionally on $\mathcal{T}_1$, $\mathcal{T}_2$ is p-complete* *w.r.t.* $\mathcal{B}_2$. Summarizing, we have proved the following Bayesian version of Theorem 2.1.

THEOREM 5.2. *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two independent statistics such that $\mathcal{B}_1 \vee \mathcal{B}_2$ is sufficient for $\mathcal{T}_1$ and $\mathcal{B}_2$ is sufficient for $\mathcal{T}_2$. If $\mathcal{T}_1 \ll \mathcal{B}_1 \mid \mathcal{B}_2$ and $\mathcal{T}_2 \ll_p \mathcal{B}_2 \mid \mathcal{T}_1$, then $(\mathcal{T}_1 \vee \mathcal{T}_2) \ll_p (\mathcal{B}_1 \vee \mathcal{B}_2)$.*

EXAMPLE 5.1. Let us illustrate Theorem 5.2 in the same discrete case as in Example 4.3. Let $(T_1, T_2, \theta_1, \theta_2) \in \{0,1\}^4$. Under both the sampling independence of T_1 and T_2, and the sufficiency of θ_2 for T_2 (see condition (4.1)(ii)), the joint probability distribution is given by

$$\omega_{ijkl} = P(T_1 = i \,|\, \theta_1 = k, \theta_2 = l) P(T_2 = j \,|\, \theta_2 = l) P(\theta_1 = k, \theta_2 = l) \doteq p_{i|kl}\, q_{j|l}\, m_{kl}. \tag{5.1}$$

Therefore, the Bayesian experiment is characterized by 9 parameters: $p_{0|00}$, $p_{0|10}, p_{0|01}, p_{0|11}; q_{0|0}, q_{0|1}$; m_{00}, m_{01}, m_{10}. Now, let us explicitly describe the matrices necessary to characterize the p-completeness relationships used in Theorem 5.2.

1. Condition $T_1 \ll_p \theta_1 \mid \theta_2$ requires to analyse the rank of the 2×2 matrices with entries of the form

$$\begin{aligned} f_{ik|l} &\doteq P[T_1 = i, \theta_1 = k \mid \theta_2 = l] \\ &= \frac{P[T_1 = i \mid \theta_1 = k, \theta_2 = l] \cdot P[\theta_1 = k, \theta_2 = l]}{P[\theta_2 = l]} = \frac{p_{i|kl} \cdot m_{kl}}{m_{+l}}, \end{aligned}$$

where $m_{+l} = m_{1l} + m_{2l}$. Therefore, for $l = 1, 2$,

$$F^{(l)} \doteq \begin{pmatrix} f_{00|l} & f_{01|l} \\ f_{10|l} & f_{11|l} \end{pmatrix} = \frac{1}{(m_{0l} + m_{1l})} \begin{pmatrix} p_{0|0l} & p_{0|1l} \\ p_{1|0l} & p_{1|1l} \end{pmatrix} \begin{pmatrix} m_{0l} & 0 \\ 0 & m_{1l} \end{pmatrix}.$$

Consequently, $T_1 \ll_p \theta_1 \mid \theta_2$ if and only if $F^{(0)}$ and $F^{(1)}$ have full rank, which is equivalent to both:

S1. $m_{kl} > 0$ for all $(k, l) \in \{0,1\}^2$;

S2. for $l = 1, 2$, $(p_{0|0l}, p_{1|0l})$ and $(p_{0|1l}, p_{1|1l})$ are linearly independent.

2. Condition $T_2 \ll_p \theta_2 \mid T_1$ requires to analyse the rank of the 2×2 matrices with entries of the form $g_{jl|i} \doteq P[T_2 = j, \theta_2 = l \mid T_1 = i]$.

Noticing that both $T_1 \perp\!\!\!\perp T_2 \mid (\theta_1, \theta_2)$ and $\theta_1 \perp\!\!\!\perp T_2 \mid \theta_2$ jointly imply that $T_1 \perp\!\!\!\perp T_2 \mid \theta_2$, it follows that

$$g_{jl|i} = \frac{q_{j|l} \cdot P[T_1 = i \mid \theta_2 = l]\, P[\theta_2 = l]}{P[T_1 = i]} = \frac{q_{j|l}\left[p_{i|0l} m_{0l} + p_{i|1l} m_{1l}\right]}{c_i},$$

where $c_i = p_{i|00} m_{00} + p_{i|01} m_{01} + p_{i|10} m_{10} + p_{i|11} m_{11}$, with $i = 0, 1$. Thus, condition $T_2 \ll_p \theta_2 \mid T_1$ relies on the following matrices.

$$G^{(i)} \doteq \begin{pmatrix} g_{00|i} & g_{01|i} \\ g_{10|i} & g_{11|i} \end{pmatrix} = c_i\, Q\, P_i\, M, \qquad \text{with } i = 1, 2,$$

where

$$Q = \begin{pmatrix} q_{0|0} & q_{0|1} \\ q_{1|0} & q_{1|1} \end{pmatrix}, \qquad M = \begin{pmatrix} m_{00} & 0 \\ m_{10} & 0 \\ 0 & m_{01} \\ 0 & m_{11} \end{pmatrix}.$$

$$P_i = \begin{pmatrix} p_{i|00} & p_{i|10} & 0 & 0 \\ 0 & 0 & p_{i|01} & p_{i|11} \end{pmatrix},$$

Therefore, $T_2 \ll_p \theta_2 \mid T_1$ if and only if $G^{(0)}$ and $G^{(1)}$ have full rank, which is equivalent to

S3. $r(M) = 2$;

S4. $r(P_i) = 2 \quad$ with $i = 1, 2$;

S5. $r(Q) = 2$.

3. Condition $(T_1, T_2) \ll_p (\theta_1, \theta_2)$ requires to analyse the rank of W as defined in (5.1). It can easily be verified that

$$W = \operatorname{diag}(Q, Q) \cdot \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} E_{32} \cdot \operatorname{diag}(m_{00}, m_{01}, m_{10}, m_{11}),$$

$$\text{with} \quad E_{32} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

It follows that (S1)–(S5) are sufficient to imply that $r(W) = 4$; indeed, condition (S5) ensures that $r[\operatorname{diag}(Q, Q)] = 4$. Condition (S1) ensures that $r[\operatorname{diag}(m_{00}, m_{01}, m_{10}, m_{11})] = 4$. Note that (S1) is more restrictive than (S3) since it implies (S3). Finally, conditions (S2) and (S4) ensure that $r[P_1' \;\; P_2'] = 4$. Let us complete this example pointing out that the implication is still valid in the case of prior independence between θ_1 and θ_2.

6 Sampling and Bayesian Versions of Joint Completeness: A Comparison

When comparing sampling and Bayesian concepts of completeness, Theorem 3.1 gives a general result of equivalence under a condition of regularity of the prior distribution when completeness is relative to the *full parameter* of a statistical model. When considered relative to a non-injective function of the parameters, the sampling theory concept is not different from completeness with respect to a full parameter, as noticed in Section 2. The situation is however different in a Bayesian framework, where completeness with respect to a non-injective function of the parameters depends on the probability measure characterizing the Bayesian experiment integrated out with respect to the prior distribution conditional on the retained parameters.

A deeper comparison between these concepts of completeness may be obtained through a comparison of their properties in specific cases. This is the object of Sections 4 and 5, where the comparison is made while combining complete statistics. Two properties are of interest: ($\boldsymbol{A}$) separate completeness of each of the two statistics $\mathcal{T}_1$ and $\mathcal{T}_2$; and ($\boldsymbol{B}$) joint completeness of $\mathcal{T}_1 \vee \mathcal{T}_2$ in the product experiment. Table 1 summarizes such differences and connections under three conditions: sampling independence of $\mathcal{T}_1$ and $\mathcal{T}_2$; variation-free between the parameters of both experiments; and non-variation-free between the parameters of both experiments. Let us remind that, in a pure sampling theory approach, variation-free means that the parameters of each experiment lie in a Cartesian product space, whereas in a Bayesian theory framework, it means that the parameters of each experiment are a priori independent. In Table 1, *L-R type theorems* means Landers and Rogge (1976) type theorem, whereas *C-K-S type theorems* means Cramer et al. (2002) type theorem.

7 Why Should a Bayesian Consider Completeness?

Let us complete this paper showing why completeness is a relevant issue in a Bayesian framework. The completeness of a statistic w.r.t. a parameter ensures that the statistic succeeds in separating the informative part of the data from the part which by itself carries no information. Basu's Second Theorem explains how a complete sufficient statistic achieves this separation, namely by making the ancillary part of the data independent of a complete sufficient statistic. Roughly speaking, a sufficient statistic is complete if it

Table 1. Sampling and Bayesian versions of L-R type theorems and C-K-S type theorems

Theorems	Set-up	$\boldsymbol{A} \Rightarrow \boldsymbol{B}$	$\boldsymbol{B} \Rightarrow \boldsymbol{A}$
L-R type	*Sampling*	Theorem 2.2, variation-free, sampling independence	Theorem 2.3, part I, variation-free, sampling independence
	Bayesian	Theorem 4.1, part I, variation-free, sampling independence	Theorem 4.1, part II, non variation-free
C-K-S type	*Sampling*	Theorem 2.1 non-variation-free, sampling independence	Theorem 2.3, part II, non-variation-free, sampling independence
	Bayesian	Theorem 5.2, non variation-free, sampling independence	Theorem 5.1, non-variation-free

contains no ancillary information or if all ancillary information is independent of it; see Lehmann (1981, pp. 337f). This explains why a complete sufficient statistic is a minimal sufficient one, so that the sampling process is fully described by it *without* redundancy.

7.1. Bayesian completeness on the parameter space and the updating Bayesian process. In this paper, we make a systematic use of the formal symmetry between parameters and observations in a Bayesian experiment. In this way, let $\mathcal{E} = (\Theta \times S, \mathcal{A} \vee \mathcal{S}, Q = m \otimes P^{\mathcal{A}})$ be a Bayesian experiment; it is said that a parameter $\mathcal{B} \subset \mathcal{A}$ is p-complete $(1 \leq p \leq \infty)$ w.r.t. a statistic $\mathcal{T} \subset \mathcal{S}$ if $\forall\, b \in L^p(\Theta \times S, \mathcal{B}, Q_{\mathcal{B} \vee \mathcal{T}})$

$$E(b \mid \mathcal{T}) = 0 \implies b = 0 \quad P_{\mathcal{T}}\text{-a.s.}, \tag{7.1}$$

where $P_{\mathcal{T}}$ is the trace of the predictive probability P on $\mathcal{T}$. We denote it as $\mathcal{B} \ll_p \mathcal{T}$. Thus, parameter completeness corresponds to the injectivity of the posterior expectation, as an operator defined on the (integrable) functions of the parameters. Similarly, a parameter $\mathcal{B}$ is p-complete w.r.t. a statistic $\mathcal{T}$ conditionally on $\mathcal{M} \subset \mathcal{A} \vee \mathcal{B}$ if and only if $\mathcal{B} \vee \mathcal{M} \ll_p \mathcal{T} \vee \mathcal{M}$. We denote it as $\mathcal{B} \ll_p \mathcal{T} \mid \mathcal{M}$.

It can be shown (see Florens et al., 1990, Theorem 5.4.12) that if $\mathcal{B}$ is a p-complete parameter w.r.t. $\mathcal{S}$, then $\mathcal{B}$ is a *minimal* sufficient parameter for $\mathcal{S}$, i.e., $\mathcal{B}$ is a sufficient parameter (see Definition 4.1), and if $\mathcal{C} \subset \mathcal{A}$ is a sufficient parameter for $\mathcal{S}$, then $\mathcal{B} \subset \overline{\mathcal{C}}$, where $\overline{\mathcal{C}} = \mathcal{C} \vee \{E \in \mathcal{A} \,:\, m(E) = m^2(E)\}$, i.e., $\overline{\mathcal{C}}$ is $\mathcal{C}$ extended so as to include all the measurable null sets; for details, see Florens et al. (1990, Chapter 2). It should be stressed that minimal sufficiency of $\mathcal{B}$ follows from the p-completeness only, *without* assuming that $\mathcal{B}$ is a sufficient parameter.

Remark 7.1. The reason for which, in a Bayesian framework, p-completeness implies minimal sufficiency can be heuristically viewed as follows: if all the information provided by a statistic is relevant for the posterior distribution (and thus the parameter is p-complete w.r.t. the statistic), it is natural that the parameter is not redundant for describing the sampling distribution.

As pointed out after Definition 4.1, a sufficient parameter $\mathcal{C} \subset \mathcal{A}$ describes the sampling process. Nevertheless, two sufficient parameters $\mathcal{C}_1$ and $\mathcal{C}_2$ provide two different descriptions of the same sampling process since $E(f \mid \mathcal{A}) = E(f \mid \mathcal{C}_1) = E(f \mid \mathcal{C}_2)$ for all $f \in L^1(S, \mathcal{S}, P)$, where P is the predictive probability. Therefore, the description of a sampling process through a sufficient parameter *always* involves redundant information. Moreover, $E(a \mid \mathcal{S} \vee \mathcal{C}_1) = E(a \mid \mathcal{C}_1)$ for all $a \in L^1(\Theta, \mathcal{A}, m)$, where m is the prior probability. Thus, the sample does not increase the prior knowledge about $\mathcal{C}_1$ given $\mathcal{A}$ or, in other words, part of the prior information on $\mathcal{A}$ is *not* revised by the sample. These considerations lead to define, in the context of the Bayesian experiment $\mathcal{E}$, parameter identification in terms of minimal sufficiency: the parameter $\mathcal{A}$ is *identified* by the observation $\mathcal{S}$ if $\mathcal{A}$ is the minimal sufficient parameter; for details, see Florens et al. (1990, Chapter 4). Thus, a parameter is said to be identified if it corresponds to the greatest possible parameter reduction for which the prior information is updated by the sample. Consequently, *an identified parameter fully concentrates the learning process underlying a Bayesian model.*

7.2. Landers and Rogge type theorem on the parameter space and its application. Parameter completeness is also useful when analysing the identifiability of hierarchical models frequently used when modelling individual data. Such an analysis is based on the following result that corresponds to a symmetrized version of Landers and Rogge's (1976) theorem; its proof is entirely similar to that of part I of Theorem 4.1.

THEOREM 7.1. *Under conditions (4.1) and (4.2)(i), if* $\mathcal{B}_i$ *is p-complete* $(1 \leq p \leq \infty)$ *w.r.t.* $\mathcal{T}_i$*, with* $i = 1, 2$*, then* $\mathcal{B}_1 \vee \mathcal{B}_2$ *is p-complete* $(1 \leq p \leq \infty)$ *w.r.t.* $\mathcal{T}_1 \vee \mathcal{T}_2$*.*

This theorem establishes that the completeness of the parameters in each marginal model imply joint completeness of the two parameters. The prior independence (4.2)(ii) is not required and, therefore, the theorem is still valid if $\mathcal{B}_i = \mathcal{D} \vee \mathcal{C}_i$ with $i = 1, 2$ and $\mathcal{D}$ different from the trivial σ-field. In particular, if $\mathcal{B}_1 = \mathcal{B}_2 \doteq \mathcal{B}$, the theorem is still valid.

SEMI-PARAMETRIC RASCH POISSON COUNTS MODELS. Let us illustrate the use of Theorem 7.1 for analysing the identifiability of a Rasch Poisson counts model (RPCM). RPCM is a unidimensional latent trait model typically used in situations, where a test consists of multiple attempts on a single item within a given time-limit. This is often the case in the assessment of psychomotor skills, where the test score is the number of successful attempts. Another situation, where the model might be appropriate, arises as a limiting case of Binomial Trials Model (Master and Wright, 1984; Jansen and Van Duijn, 1991); there the test score is the number of incorrect responses out of m items with low, but not necessarily equal, error probabilities. The conditional probability of observing the response Y_{ij} of subject $i = 1, \ldots, n$ on test $j = 1, \ldots, m$ is given by

$$(Y_{ij} \mid \lambda_{ij}) \sim \text{Poisson}(\lambda_{ij}), \quad \lambda_{ij} = \exp(\theta_i - \beta_j), \tag{7.2}$$

where $\beta_j \in \mathbb{R}$ is the difficulty of test j and $\theta_i \in \mathbb{R}$ is the ability of person i; here (θ_i, β_j) is a sufficient parameter for Y_{ij} in the sense of Definition 4.1, but it is not a minimal sufficient parameter. The model specification is completed with the following structural hypotheses.

- H1. $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ are mutually independent given $(\theta_1, \ldots, \theta_n, \beta_1, \ldots, \beta_m)$, where $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im})' \in \mathbb{N}^m$.
- H2. For each person i, the distribution of $\boldsymbol{Y}_i$ depends on $(\theta_i, \beta_1, \ldots, \beta_m)$ only.
- H3. For each person i, his/her responses $\{Y_{ij} : 1 \leq j \leq m\}$ are mutually independent given $(\theta_i, \beta_1, \ldots, \beta_m)$; this corresponds to the Axiom of Local Independence.

H4. The abilities $\theta_1, \ldots, \theta_n$ are mutually independent given G, where G is a random probability measure defined on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Moreover, $(\theta_i \mid G) \sim G$ for all i.

The inference is based on the *statistical model* obtained after integrating out the unobserved abilities θ_i's. The structural hypotheses H1 to H4 ensure that the responses $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ are mutually independent given $(\beta_1, \ldots, \beta_m, G)$ with a common probability distribution given by

$$P[\boldsymbol{Y}_1 = \boldsymbol{\epsilon}_1 \mid \beta_1, \ldots, \beta_m, G] \;=\; \int_{\mathbb{R}} \prod_{1 \leq j \leq m} P[Y_{1j} = \epsilon_{1j} \mid \theta, \beta_j]\, G(d\theta),$$

where $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{im})' \in \mathbb{N}^m$. Thus, the parameters of interest are the difficulty of the m tests $\beta_1, \ldots, \beta_m$ and the probability distribution G generating the abilities. In this case, the parameter space is given by the product space $\mathbb{R}^m \times \mathcal{P}(\mathbb{R}, \mathcal{B})$.

Identifiability of a semi-parametric Rasch Poisson counts model. The problem is to establish restrictions under which $(\beta_1, \ldots, \beta_m, G)$ is the minimal sufficient parameter for $\boldsymbol{Y}_1$, and so identified by it. Using Theorem 1 of Mouchart and San Martín (2003), it can be proved that $(\beta_1, \ldots, \beta_m, G)$ is identified by $\boldsymbol{Y}_1$ conditionally on β_1 provided the following two conditions are satisfied.

(1) θ_1 is identified by G conditionally on β_1.

(2) $(\beta_1, \ldots, \beta_m, \theta_1)$ is 2-complete w.r.t. $\boldsymbol{Y}_1$ conditionally on β_1.

Condition (1) is straightforward since $P\{\theta_1 \in (-\infty, x] \mid G\} = G(x)$ for all $x \in \mathbb{R}$, so G is identified by θ_1. But by H4, $\theta_1 \perp\!\!\!\perp \beta_1 \mid G$; therefore, G is identified by θ_1 conditionally on β_1. Condition (2) above is obtained using the Landers and Rogge type Theorem 7.1. As a matter of fact, it can be proved that if $(X \mid \lambda) \sim \text{Poisson}(\lambda)$, then $\lambda \ll_2 X$ for all prior distribution μ_λ defined on $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$; for a proof, see San Martín (2000). Since $\exp(\cdot)$ is a bijective function and $(Y_{ij} \mid \lambda_i, \beta_j) \sim \text{Poisson}(\exp(\theta_i - \beta_j))$, it follows that

$$\theta_1 - \beta_j \ll_2 Y_{1j} \quad \forall\, 1 \leq j \leq m. \tag{7.3}$$

For $m \geq 2$, Theorem 7.1 ensures that, under the hypothesis H3, condition (7.3) implies that $(\theta_1 - \beta_1, \ldots, \theta_1 - \beta_m) \ll_2 Y_1 \mid \beta_1$, which is equivalent to $(\beta_1, \ldots, \beta_m, \theta_1) \ll_2 Y_1 \mid \beta_1$.

Thus, $(\beta_1, \ldots, \beta_m, G)$ is identified by $\boldsymbol{Y}_1$ conditionally on β_1, which in practice means to fix the difficulty of the first test at 0.

Appendix: Proofs

Proof of Theorem 4.1 is based on the following general result established in Florens et al. (1990).

THEOREM A.1. *Let* $p \in [1, \infty]$ *and let* X_1, X_2, X_3, X_4 *be random variables defined on a common probability space* $(\Omega, \mathcal{M}, P)$. *If* $X_2 \perp\!\!\!\perp X_4 \mid (X_1, X_3)$ *then*

(i) $X_1 \ll_p X_2 \mid X_3 \Rightarrow X_1 \ll_p X_2 \mid (X_3, X_4)$ *(see Florens et al., 1990, Theorem 5.4.5).*

(ii) $X_2 \ll_p (X_1, X_4) \mid X_3 \Rightarrow X_2 \ll_p X_1 \mid X_3$ *(see Florens et al., 1990, Theorem 5.4.6).*

PROOF OF THEOREM 4.1, PART I. Conditions (4.1) and (4.2) imply that $\mathcal{T}_2 \perp\!\!\!\perp \mathcal{B}_1 \mid \mathcal{T}_1$. This condition, along with the p-completeness of $\mathcal{T}_1$ w.r.t. $\mathcal{B}_1$, imply, by part (i) of Theorem A.1, that **(a)** $(\mathcal{T}_1 \vee \mathcal{T}_2) \ll_p (\mathcal{B}_1 \vee \mathcal{T}_2)$. Similarly, conditions (4.1)(ii) and (4.2)(ii) imply that $\mathcal{B}_1 \perp\!\!\!\perp \mathcal{B}_2 \mid \mathcal{T}_2$. This last condition, along with the p-completeness of $\mathcal{T}_2$ w.r.t. $\mathcal{B}_2$, jointly imply that **(b)** $(\mathcal{B}_1 \vee \mathcal{T}_2) \ll_p (\mathcal{B}_1 \vee \mathcal{B}_2)$. Let $f \in L^p(M, \mathcal{T}_1 \vee \mathcal{T}_2, Q)$. Since $\mathcal{T}_1 \vee \mathcal{T}_2 \perp\!\!\!\perp \mathcal{B}_1 \vee \mathcal{B}_2 \mid \mathcal{B}_1 \vee \mathcal{T}_2$ (a property implied by (4.1) and (4.2)), it follows that $E(f \mid \mathcal{B}_1 \vee \mathcal{B}_2) = E[E(f \mid \mathcal{B}_1 \vee \mathcal{T}_2) \mid \mathcal{B}_1 \vee \mathcal{B}_2]$. Therefore, if $E(f \mid \mathcal{B}_1 \vee \mathcal{B}_2) = 0$, property **(b)** above implies that $E(f \mid \mathcal{B}_1 \vee \mathcal{T}_2) = 0$; and, by property **(a)** above, $f = 0$ Q-a.s. $\square$

PROOF OF THEOREM 4.1, PART II. If $(\mathcal{T}_1 \vee \mathcal{T}_2) \ll_p (\mathcal{B}_1 \vee \mathcal{B}_2)$, then $\mathcal{T}_1 \ll_p (\mathcal{B}_1 \vee \mathcal{B}_2)$ since the set of $\mathcal{T}_1$-measurable functions is contained in the set of $(\mathcal{T}_1 \vee \mathcal{T}_2)$-measurable functions. This condition along with (4.1)(i) imply, by part (ii) of Theorem A.1, that $\mathcal{T}_1 \ll_p \mathcal{B}_1$. $\square$

PROOF OF PROPOSITION 4.1. By Definition 3.1, $X_1 \ll_p X_2 \mid X_3$ if and only if $(X_1, X_3) \ll_p (X_2, X_3)$. Therefore, the proposition needs to be proved for $X_3 = E(X_3)$ a.s. The equivalence (ii) $\Longleftrightarrow$ (iii) is a consequence of the rank theorem (see Halmos, 1974, Theorem 1, Section 50), whereas the equivalence of (iii) $\Longleftrightarrow$ (iv) follows from Lemma 4.1 in San Martín et al. (2005).

The implication (i) $\Longrightarrow$ (iii) is proved as follows. Let $d \in \mathrm{Ker}[C(X_2, X_1)]$. Then, under normality, it follows that $E[f(d'X_1) \mid X_2] = E[f(d'X_1)]$ for all

$f \in L^p(\mathbb{R}^{p_1}, \mathcal{B}_{\mathbb{R}^{p_1}}, P)$, where $\mathcal{B}_{\mathbb{R}^{p_1}}$ denotes the Borel sets of $\mathbb{R}^{p_1}$. This is equivalent to

$$E\{f(d'X_1) - E[f(d'X_1)] \mid X_2\} = 0 \quad \text{a.s. } \forall\, f \in L^p(\mathbb{R}^{p_1}, \mathcal{B}_{\mathbb{R}^{p_1}}, P). \qquad \text{(A.1)}$$

Since $X_1 \ll_p X_2$, equality (A.1) implies that $f(d'X_1) = E[f(d'X_1)]$ a.s. for all $f \in L^p(\mathbb{R}^{p_1}, \mathcal{B}_{\mathbb{R}^{p_1}}, P)$. It follows that $V(d'X_1) = 0$, so that $\mathrm{Ker}[C(X_2, X_1)] \subset \mathrm{Ker}[V(X_1)]$. The other inclusion follows from Lemma 4.1 in San Martín et al. (2005).

The implication (iv) $\Longrightarrow$ (i) follows from the following steps.

1. Using Lemma C.1 in San Martín et al. (2005), there exists a matrix A_2 such that $X_2^* = A_2'X_2 \in \mathbb{R}^{q_2}$, where $q_2 = r[V(X_2)]$, $X_2^* = X_2$ a.s., $V(X_2^*)$ is positive definite and $(X_1 \mid X_2^*) \sim \mathcal{N}_{p_1}(g + R_{12}^* X_2^*, V(X_1 \mid X_2^*))$, with $R_{12}^* = C(X_1, X_2^*)V(X_2^*)^{-1}$. Consequently, (iv) is rewritten as (iv′): $\mathrm{Ker}\,(R_{12}^{*\prime}) \subset \mathrm{Ker}[V(X_1 \mid X_2^*)]$; whereas (i) is rewritten as (i′): $X_1 \ll_p X_2^*$.

2. Let $q_1 = r[V(X_1 \mid X_2*)] \leq p_1$. By a similar argument, there exists a matrix A_1 of dimension $p_1 \times q_1$ and a matrix C_1 of dimension $p_1 \times (p_1 - q_1)$ such that $r(A_1) = q_1$, $r(C_1) = p_1 - q_1$, $A_1'C_1 = 0$, $\mathrm{Im}\,[V(X_1 \mid X_2^*)] = \mathrm{Im}\,(A_1)$, $\mathrm{Ker}\,[V(X_1 \mid X_2^*)] = \mathrm{Im}\,(C_1)$ and $V(A_1'X_1 \mid X_2^*)$ is a positive definite matrix.

3. Let $s = r(R_{12}^*) \leq \min\{p_1, q_2\}$. From the singular value decomposition of R_{12}^*, there exist two orthonormal matrices A_4 and A_5 such that $R_{12}^* = A_4 \Delta A_5'$, where $\mathrm{Im}\,(R_{12}^*) = \mathrm{Im}\,(A_4)$, $\mathrm{Ker}\,(R_{12}^*) = \mathrm{Ker}\,(A_5')$, $r(A_4) = r(A_5) = s$, and Δ is a diagonal positive definite matrix. Thus, using steps (1) and (2), condition (iv′) is equivalently to $\mathrm{Im}\,(A_1) \subset \mathrm{Im}\,(A_4)$, which in turn implies that $q_1 \leq s$.

4. Thus, if $\mathrm{Im}\,(A_1) \subset \mathrm{Im}\,(A_4)$, we can take $A_4 = (A_1 \quad G_1)$, where G_1 is a $p_1 \times (s - q_1)$ matrix, and $C_1 = (G_1 \quad C_4)$, where C_4 is such that $Q_4 = (A_4 \quad C_4)$ is a $p_1 \times p_1$ orthonormal matrix. Now let

$$\begin{aligned}
&\text{(i)} \quad Z &&= Q_4{}'X_1 = (A_1, G_1, C_4)X_1 \\
& &&= (Z_1{}', Z_2{}', Z_3{}')' \in \mathbb{R}^{q_1} \times \mathbb{R}^{s-q_1} \times \mathbb{R}^{p_1-s}, \\
&\text{(ii)} \quad V &&= \begin{pmatrix} \Delta & 0 \\ 0 & I_{q_2-s} \end{pmatrix} \begin{pmatrix} A_5{}' \\ C_5{}' \end{pmatrix} X_2^* \\
& &&= (V_1{}', V_2{}', V_3{}')' \in \mathbb{R}^{q_1} \times \mathbb{R}^{s-q_1} \times \mathbb{R}^{p_2-s},
\end{aligned} \qquad \text{(A.2)}$$

where C_5 is such that $(A_5 \quad C_5)$ is a $q_2 \times q_2$ orthonormal matrix. It follows that $X_1 = Z$ a.s. and $X_2^* = V$ a.s.. Consequently, $X_1 \ll_1 X_2^*$ is equivalent to $(Z_1, Z_2, Z_3) \ll_1 (V_1, V_2, V_3)$. Moreover, from step 3 and (A.2)(ii), it follows that $R_{12}^* X_2^* = (A_4 \quad 0)V$. Thus, since $X_2^* = V$ a.s.,

$$(X_1 \mid V_1, V_2, V_3) \sim \mathcal{N}_{p_1}(g + A_1 V_1 + G_1 V_2,\ V(X_1 \mid X_2^*)).$$

But, using (A.2)(i), it follows that $Z_3 = {C_4}' g$ a.s., $Z_2 = G_1' g + V_2$ a.s. and

$$(Z_1 \mid V_1, V_2, V_3) \sim \mathcal{N}_{q_1}({A_1}' g + V_1,\ V({A_1}' X_1 \mid X_2^*)).$$

These relations imply that $(Z_1, Z_2, Z_3) \ll_1 (V_1, V_2, V_3)$, which is equivalent to $Z_1 \ll_1 (V_1, V_3) \mid V_2$; and $\mathcal{Z}_1 \perp\!\!\!\perp \mathcal{V}_2 \vee \mathcal{V}_3 \mid \mathcal{V}_1$. Under this last conditional independence, $Z_1 \ll_1 (V_1, V_3) \mid V_2$ is equivalent to $Z_1 \ll_1 V_1 \mid V_2$. Since $V({A_1}' X_1 \mid X_2^*) > 0$, by fixing V_2 in the conditional distribution of $(Z_1 \mid V_1)$, the proof follows by using the fact that Z_1 is a complete statistic in $L^1(\mathbb{R}^{p_1}, \mathcal{B}_{\mathbb{R}^{p_1}}, P)$ with respect to V_1; see Barndorff-Nielsen (1978, Lemma 8.2). Consequently, condition (iv) implies that $X_1 \ll_1 X_2$. Since $L^p(\mathbb{R}^{p_1}, \mathcal{B}_{\mathbb{R}^{p_1}}, P) \subset L^1(\mathbb{R}^{p_1}, \mathcal{B}_{\mathbb{R}^{p_1}}, P)$, it follows that $X_1 \ll_p X_2$. $\square$

PROOF OF PROPOSITION 4.2: By Definition 3.1, it is enough to prove this proposition for $X_3 = E(X_3)$ a.s. Let $N_2^* = \{j \in N_2 : P(X_2 = j) > 0\}$, and for $(i, j) \in N_1 \times N_2^*$, consider the $|N_1| \times |N_2^*|$ matrix $\mathbf{P}_{1|2} = [(P[X_1 = i \mid X_2 = j])_{ij}]$. It follows that, for $j \in N_2^*$, $E[g(X_1) \mid X_2 = j] = g' \mathbf{P}_{1|2}\, e_j$, where e_j is the j-th column of $I_{|N_2^*|}$. Then X_1 is p-complete w.r.t. X_2 if and only if the following implication holds: $g' \mathbf{P}_{1|2} = 0 \implies g = 0$; that is, if $\mathbf{P}'_{1|2}$ is an injective linear transformation, or, equivalently, if $\mathbf{P}'$ is an injective linear transformation. $\square$

Acknowledgements. The first author gratefully acknowledges the financial support of the FONDECYT Project N°1060722 from the Corporación Nacional de Ciencia y Tecnología, Chile. The second author acknowledges the financial support from the IAP research network P6/03 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs). Both authors gratefully acknowledge a partial financial support from the Measurement Center MIDE UC (Chile) to complete this paper. The last revision of the paper was done while the first author visited the Faculty of Psychology, K.U. Leuven. Both authors are grateful to the Co-editor and the referee for their comments leading to improvement of the paper.

References

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory.* Wiley, New York.

Barra, J.-R. (1981). *Mathematical Basis of Statistics.* Academic Press, New York.

Basu, D. (1955). On statistics independent of a complete sufficient statistics. *Sankhyā*, **15**, 377–380.

Basu, D. (1959). The family of ancillary statistics. *Sankhyā*, **21**, 247–256.

Basu, D. (1964). Recovery of ancillary statistics. *Sankhyā Ser. A*, **26**, 3–16.

Basu, D. (1977). On the elimination of nuisance parameters. Part I. *J. Amer. Statist. Assoc.*, **72**, 355–366.

Basu, D. and Pereira, C. B. (1983). Conditional independence in statistics. *Sankhyā Ser. A*, **45**, 324–337.

Berti, P., Fattorini, L. and Rigo, P. (2000). Eliminating nuisance parameters: Two characterizations. *Test*, **9**, 133–148.

Cramer, E., Kamps, U. and Schenk, N. (2002). On the joint completeness of independent statistics. *Statist. Decisions*, **20**, 269–277.

Dawid, P.A. (1980). Conditional independence for statistical operations. *Ann. Statist.*, **8**, 598–617.

Döhler, R. (1980). On the conditional independence of random events. *Theory Probab. Appl.*, **25**, 628–634.

Florens, J.P. and Rolin, J.-M. (1984). Asymptotic sufficiency and exact estimability in Bayesian experiments. In *Alternative Approaches to Time Series Analysis*, J.P. Florens, M. Mouchart, J.-P. Raoult and L. Simar (eds.), Publication des Facultés Universitaires Saint Louis, Bruxelles, 121–142.

Florens, J.-P., Mouchart, M. and Rolin, J.-M. (1990). *Elements of Bayesian Statistics.* Marcel Dekker, New York.

Halmos, P. (1974). *Finite-Dimensional Vector Space.* Springer, New York.

Jansen, M.G.H. (1994). Parameters of the latent distribution in Rasch's Poisson counts model. In *Contribution to Mathematical Psychology, Psychometrics, and Methodology,* G.H. Fischer and D. Laming (eds.), Springer, New York, 319–326.

Jansen, M.G.H. and van Dujin, M.A.J. (1992). Extensions of Rasch's multiplicative Poisson model. *Psychometrika*, **57**, 405-414.

Landers, D. and Rogge, L. (1976). A note on completeness. *Scandinavian J. Statist.*, **3**, 139.

Lehmann, E. L. (1981). An Interpretation of completeness and Basu's theorem. *J. Amer. Statist. Assoc.*, **76**, 335–340.

Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd. edition. Springer, New York.

Lehmann, E.L. and Scheffé, H. (1950). Completeness, similar regions and unbiased tests, Part I. *Sankhyā*, **10**, 305–340.

Lehmann, E.L. and Scheffé, H. (1955). Completeness, similar regions and unbiased tests, Part II. *Sankhyā*, **15**, 219–236.

Marsaglia, G. (1964). Conditional means and covariances of normal variables with singular covariance matrix. *J. Amer. Statist. Assoc.*, **59**, 1203–1204.

Master, G.N. and Wright, B.D. (1984). The essential process in a family of measurement models. *Psychometrika*, **49**, 529–544.

Martin, F., Petit, J.-L. and Littaye, M. (1973). Indépendance conditionelle dans le modèle statistique bayésien. *Ann. Inst. H. Poincaré Sect. B*, **9**, 19–40.

McCullagh, P. (2002). What is a statistical model? (with discussion). *Ann. Statist.*, **30**, 1225–1310.

Mouchart, M. and Rolin, J.-M. (1984). A note on conditional independence with statistical applications. *Statistica*, **44**, 557–584.

Mouchart, M. and San Martín, E. (2003). Specification and identification issues in models involving a latent hierarchical structure. *J. Statist. Plann. Inf.*, **111**, 143–163.

San Martín, E. (2000). *Latent Structural Models: Specification and Identification Problems.* Ph. D. Dissertation, Institute of Statistics, Université Catholique de Louvain, Belgium.

San Martín, E., Mouchart, M. and Rolin, J.-M. (2005). Ignorable common information, null sets and Basu's first theorem. *Sankhyā*, **67**, 674–698.

Ernesto San Martín
Department of Statistics, and
Measurement Center MIDE UC
Pontificia Universidad Católica de Chile
Casilla 306, Santiago 22, Chile
E-mail: esanmart@mat.puc.cl

Michel Mouchart
Institut de Statistique
Université catholique de Louvain
20 Voie du Roman Pays
B-1348 Louvain-la-Neuve, Belgium
E-mail: michel.mouchart@uclouvain.be

Paper received September 2006; revised December 2007.